Generating Stylistically Consistent Dialog Responses with Transfer Learning

Reina Akama[1]

Kazuaki Inada[1], Naoya Inoue[1], Sosuke Kobayashi[2], Kentaro Inui[1,3]

[1] Graduate School of Information Sciences, Tohoku University

[2] Preferred Networks, Inc.

[3] RIKEN Center for Advanced Intelligence Project

Prior approaches

- Feed large-scale conversational data into sequence-to-sequence (seq2seq) model [Vinyals+'15, Li+'16, etc.]
- Problem: generates **NOT stylistically consistent responses**

Generating Stylistically Consistent Dialog Responses
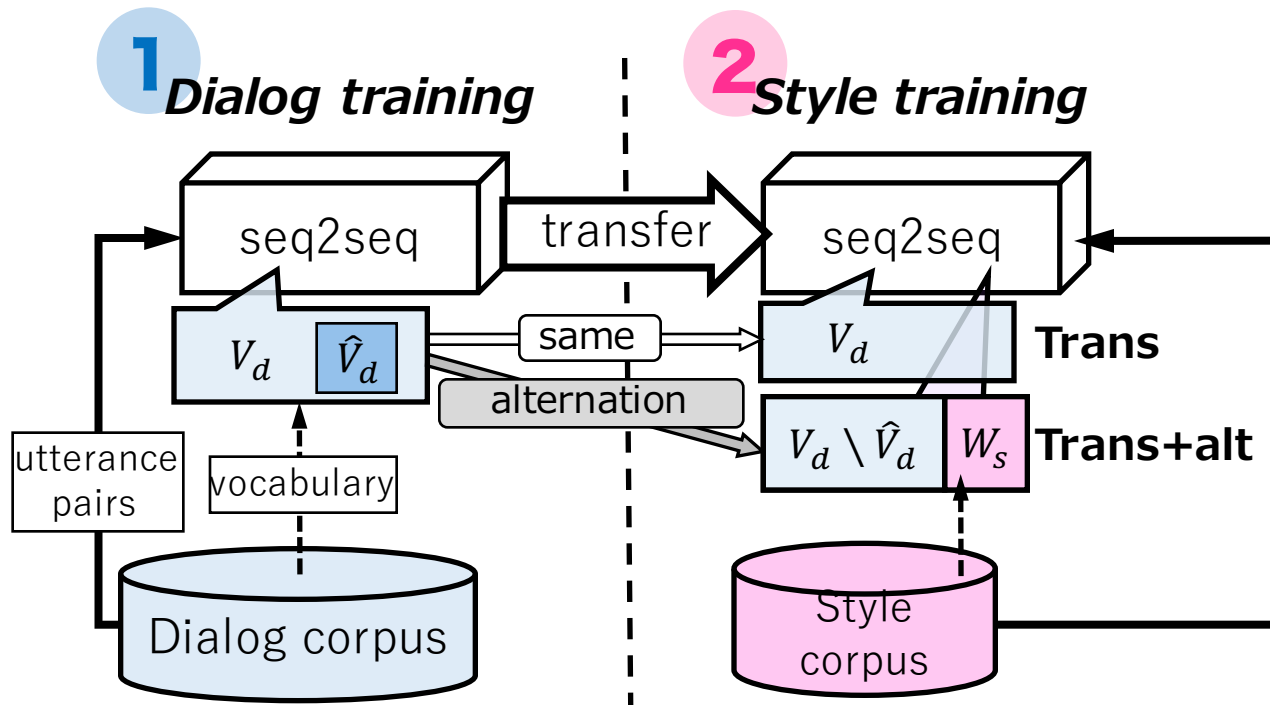with Transfer Learning
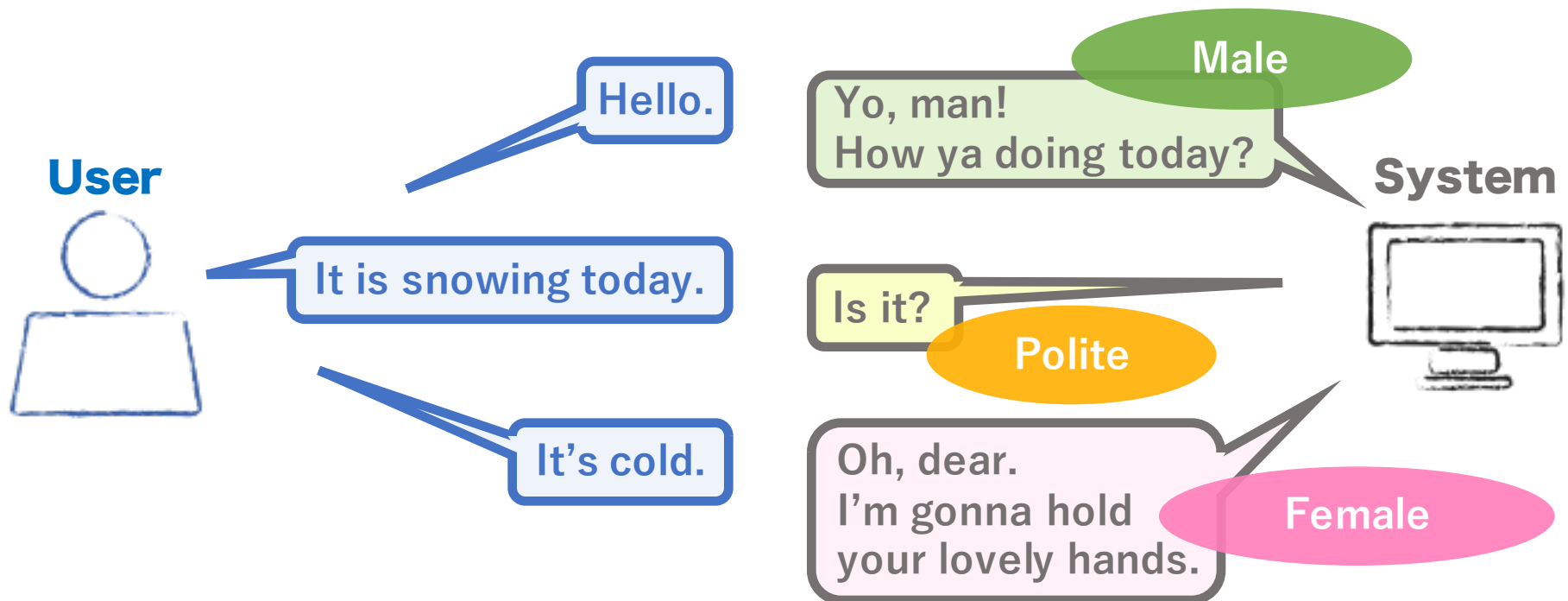
Our goal

- Generate dialog responses with **particular style consistently** while maintaining goodness of contents and fluency
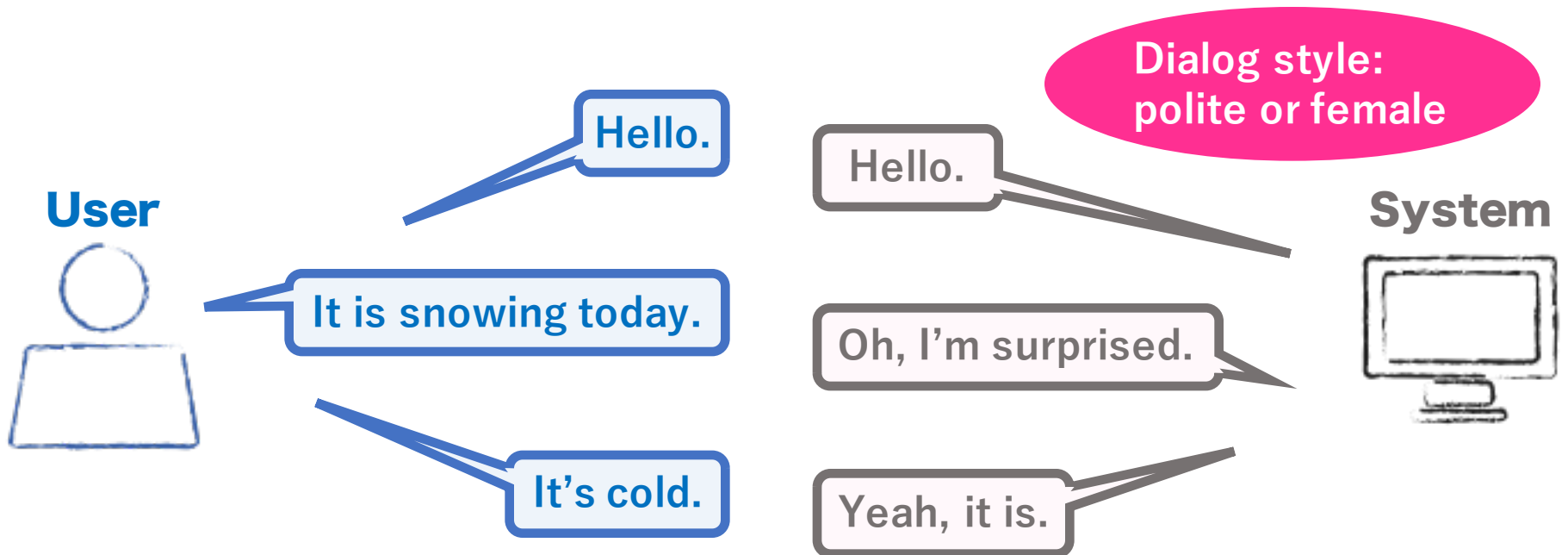
Proposed method

- Apply **transfer learning** to seq2seq response generation model
- Propose **two-staged** training framework

Generating Stylistically Consistent Dialog Responses
with Transfer Learning

- Focus on learning **what-to-say** and **fluent** responses
- Exploit cheap, large-scale but *stylistically inconsistent* dialog corpus

Generating Stylistically Consistent Dialog Responses
with Transfer Learning
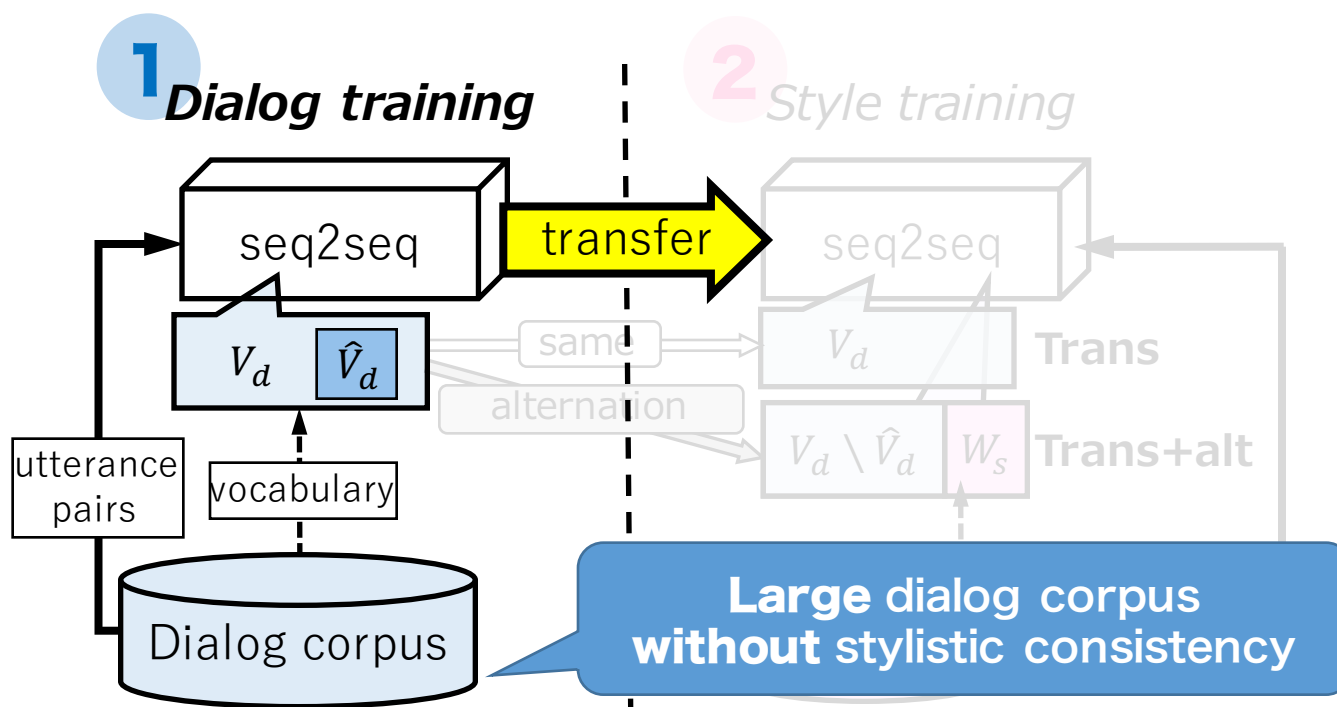
- Focus on learning **stylistically consistent** responses
- Fine-tune seq2seq model trained in dialog training
- Exploit expensive, small-scale but *stylistically consistent* corpus
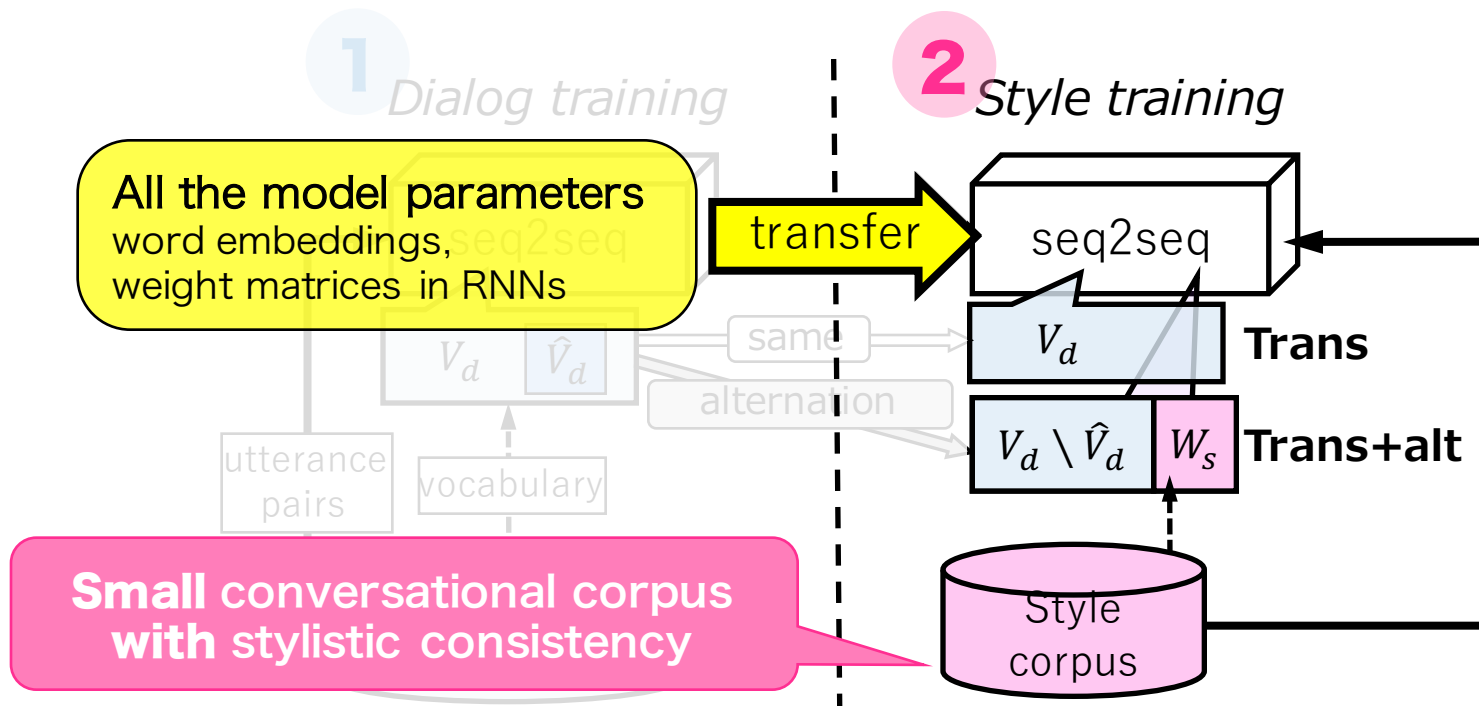
Issue: how to create the seq2seq vocab. for style training?

1. Trans

- Simply use the same vocabulary V_d as Dialog training
- V_d: the top N_d most frequent words in dialog corpus
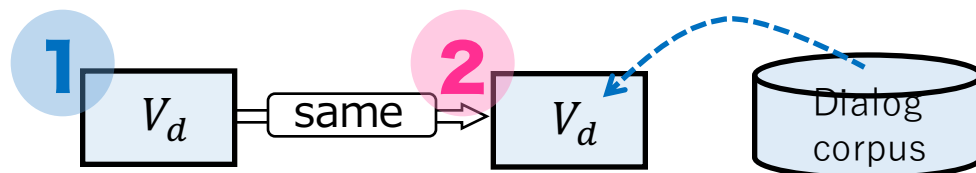
2. Trans+alt

- Alter the vocabulary before style training
- Replace infrequent words $\hat{V}_d$ from dialog corpus with frequent words W_s in style corpus

Generating Stylistically Consistent Dialog Responses
with Transfer Learning

Dataset

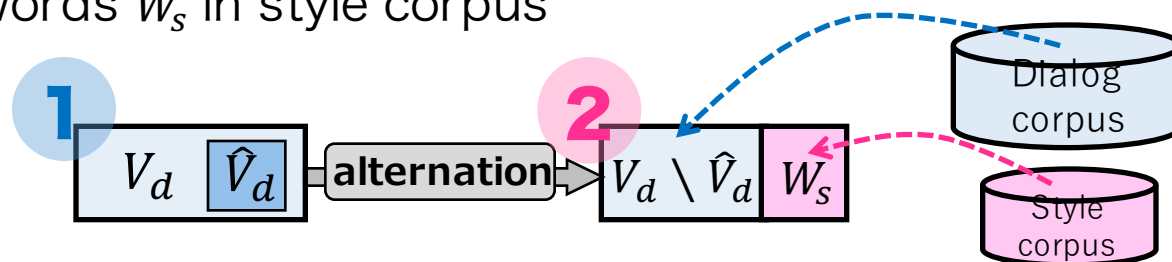- Japanese single-turn dialog (i.e., utterance pairs)

- **Dialog corpus**
  - **3.7 M** utterance pairs extracted from tweet-reply chains
- **Style corpus**
  - **Feminine and polite style**
  - **12 K** utterance pairs extracted from TV subtitles
    - Only 0.3% of dialog corpus
  - TV program where *Tetsuko Kuroyanagi* chats with some guests

  > ∗ **Japanese TV personality**
  > ∗ **Elderly woman**
  > ∗ **Speaks** feminine and polite words

Experiments setup

Models

features \ models	Baselines		Proposals	
	Base	Mix	Trans	Trans+alt
Dialog corpus	✔	✔	✔	✔
Style corpus		✔	✔	✔
Vocabulary alternation		✔		✔
Transfer learning			✔	✔

Settings

- Seq2seq: 2 layer LSTMs with 2048 units
- Optimization: Adam (mini-batch size 64)
- Word embeddings: 1024 dim
- Vocab. size: 25K

Vocabulary alternation (Trans+alt)

- Vocab. size: 1K words from style corpus (W_s)

$$V_d \quad \hat{V}_d$$

25K

alternation

$$V_d \setminus \hat{V}_d \quad W_s$$

24K + **1K**

Generating Stylistically Consistent Dialog Responses
with Transfer Learning
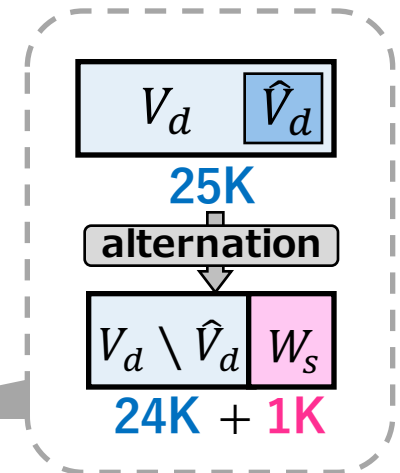
Evaluation setup

Human evaluation

- 5 crowd workers judged each responses generated by each model
- Given input, generated response, style description,
  and **2 questions:**
  Q1. Whether the response is grammatically and semantically appropriate
  Q2. Whether the style of response matches the given description

Q1: OK

input:	generated:	style:
It is snowing today.	**Oh, I'm surprised.**	**feminine, polite**

Q2: OK

Evaluation metrics

- The percentage of responses judged as
  1. appropriate responses
  2. stylistically consistent

Results
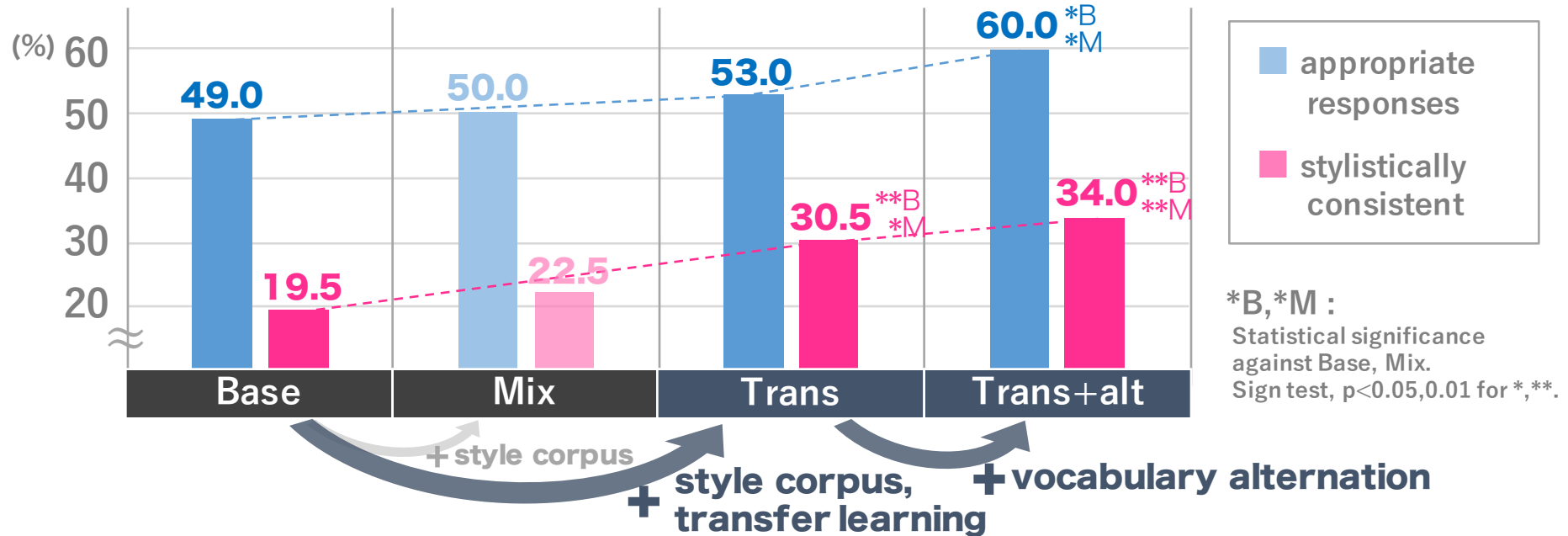
- **Transfer learning** frameworks are successful
  - Increase stylistic consistency of generated responses
  - Maintain the appropriateness of responses

- **Vocabulary alternation** helps to make the more stylistically consistent responses

Generating Stylistically Consistent Dialog Responses
with Transfer Learning

Conclusion

Summary

- Presented novel end-to-end framework to build **stylistically consistent** dialog response generation system
- Apply transfer learning to seq2seq response generation model
- Human evaluation demonstrated that proposed method produces stylistically consistent responses while maintaining appropriateness of responses

Future work

- Improve style training so that it can learn only the style of responses
- Explore effective way of creating style corpus
  - e.g., automatically collecting polite utterance from a large Twitter corpus

**Generating Stylistically Consistent Dialog Responses
with Transfer Learning**

Appendixes

Generating Stylistically Consistent Dialog Responses with Transfer Learning

References

- Diederik Kingma and Jimmy Ba. 2015. "Adam: A method for stochastic optimization." In *The Inter- national Conference on Learning Representations (ICLR)*.

- Ilya Sutskever, Oriol Vinyals, and Quoc V. Le. 2014. "Sequence to sequence learning with neural networks." In *Advances in neural information processing systems*, pages 3104–3112.

- Oriol Vinyals and Quoc Le. 2015. "A neural conversational model." In *International Conference on Machine Learning(ICML) Deep Learning Workshop 2015*.

Generated responses

- Generated responses by out best model Trans+alt
- Feminine and polite style

input	generated responses
お住まいはどちらでしょうか？ **Where do you live?**	京都です **I live in Kyoto.**
ドル安ですね **The dollar is weak, isn't it?**	そうですね **Yes, it is.**
月が綺麗 **The moon is beautiful.**	すごい **Marvelous!**

Human evaluation

- Each model generated 200 responses for evaluation
  - input: 50 utterances
  - generated 4 responses for each utterance

- Each response is judged by 5 crowd workers
  - Via Yahoo! Crowd Sourcing (https://crowdsourcing.yahoo.co.jp/)
  - Given input, generated response, **style description** and 2 yes-no questions
  - Final answer is determined by majority vote

> ∗ **only a description with several example**
> ∗ **without the specific name of individual character**